Asymptotic Analysis of Complex LASSO via Complex Approximate Message Passing (CAMP)

Arian Maleki[*], Laura Anitori[†], Zai Yang [‡], and Richard G. Baraniuk[*]

August 3, 2011

Abstract

Recovering a sparse signal from an undersampled set of random linear measurements is the main problem of interest in compressed sensing. In this paper, we consider the case where both the signal and the measurements are complex. We study the popular reconstruction method of ℓ_1-regularized least squares or LASSO. While several studies have shown that the LASSO algorithm offers desirable solutions under certain conditions, the precise asymptotic performance of this algorithm in the complex setting is not yet known. In this paper, we extend the approximate message passing (AMP) algorithm to the complex signals and measurements and obtain the complex approximate message passing algorithm (CAMP). We then generalize the state evolution framework recently introduced for the analysis of AMP, to the complex setting. Using the state evolution, we derive accurate formulas for the phase transition and noise sensitivity of both LASSO and CAMP.

1 Introduction

Recovering a sparse signal from an undersampled set of random linear measurements is a problem of interest in compressed sensing (CS). In the past few years many algorithms have been proposed for signal recovery and their performances have been analyzed both analytically and empirically [1–6]. However, whereas most of the theoretical work has focussed on the case of real-valued signals and measurements, in many applications, such as magnetic resonance imaging and radar, the signals are more easily representable in the complex domain. For such applications, often the real and imaginary components of a complex signal tend to be either zero or non-zero simultaneously. Therefore, recovery algorithms may benefit from this prior knowledge. Indeed the results presented in this paper confirm this intuition.

Motivated by these observations, we investigate the performance of the complex valued LASSO in the case of noise-free and noisy measurements. The derivations are based on the state evolution (SE) framework, presented previously in [3]. Also a new algorithm, complex approximate message passing (CAMP), is presented to solve the complex LASSO. This algorithm is an extension of the AMP algorithm [3, 7]. However, the extension of AMP and its analysis from the real to the complex setting is not trivial, due to the different properties of the amplitude and phase components compared to the real valued case.

In the follow up of this section, we briefly review some of the existing algorithms for sparse signal recovery in the real-valued setting and then focus on recovery algorithms for the complex case, with particular attention to the AMP and CAMP algorithms. We then introduce two criteria which are used

[*]Department of Electrical and Computer Engineering, Rice University

[†]TNO, The Hague, The Netherlands

[‡]Department of Electrical and Electronic Engineering, Nanyang Technological University

as measures of performance for the noiseless and noisy measurements respectively. Based on these criteria we establish the novelty of our results compared to the existing work. An overview of the organization of the paper is provided in Section 1.7.

1.1 Real-valued sparse recovery algorithms

Consider the problem of recovering a sparse vector $s_o \in \mathbb{R}^N$ from a noisy undersampled set of linear measurements $y \in \mathbb{R}^n$, where $y = As_o + w$ and w is the noise. Let k denote the number of nonzero elements of s_o. The measurement matrix has iid elements from a given distribution on $\mathbb{R}$. Given y and A, we seek an approximation to s_o. Many recovery algorithms have been proposed, ranging from the convex relaxation techniques to greedy approaches to iterative thresholding schemes. See [1] and the references therein for an exhaustive list of algorithms. [6] has compared several different recovery algorithms and concluded that among the algorithms compared in that paper the ℓ_1-regularized least squares , a.k.a. LASSO or BPDN, [2,8] that seeks the minimizer of $\min_x \frac{1}{2}\|y-Ax\|_2^2+\lambda\|x\|_1$ provides the best performance in the sense of sparsity measurement tradeoff. Recently, several iterative thresholding algorithms have been proposed for solving LASSO using few computations per iteration that enable the usage of LASSO for high dimensional problems. See [9] and the references therein for an exhaustive list of the algorithms. In this paper, we are particularly interested in AMP [3]. Starting from $x^0 = 0$ and $z^0 = y$, AMP uses the following iterations:

$$\begin{aligned} x^{t+1} &= \eta_\circ\big(x^t + A^T z^t; \tau_t\big), \\ z^t &= y - Ax^t + \frac{|I^t|}{n} z^{t-1}, \end{aligned}$$

where $\eta_\circ(x;\tau) = (|x| - \tau)_+ \text{sign}(x)$ is the soft thresholding function, τ_t is the threshold parameter and I^t is the active set of x^t, i.e., $I^t = \{i \mid x_i^t \neq 0\}$. Furthermore, the theoretical prediction of its asymptotic behavior, is also accurate for LASSO [7], [10].

1.2 Complex-valued sparse recovery algorithms

Consider the complex setting where both the signal s_o and the measurements y are in the complex domain. The success of LASSO has motivated researchers to use similar techniques here as well. We consider the following two schemes that have been used in the signal processing literature:

- r-LASSO -- The simplest extension of the LASSO to the complex setting is to consider the complex signal and measurements as a $2N$ dimensional real signal and $2n$ dimensional real measurements, respectively. Let the superscript R and I denote the real and imaginary parts of a complex number. Define $\tilde{y} = [(y^R)^T, (y^I)^T]^T$ and $\tilde{s_o} = [(s_o^R)^T, (s_o^I)^T]^T$, where the superscript T denotes the transpose operator. We have

$$\tilde{y} = \underbrace{\begin{pmatrix} A^R & -A^I \\ A^I & A^R \end{pmatrix}}_{\tilde{A}} \tilde{s}_o.$$

  We then search for an approximation of $\tilde{s}_o$ by running $\arg\min_x \frac{1}{2}\|\tilde{y} - \tilde{A}x\|_2^2 + \lambda\|x\|_1$ [11,12]. We call this algorithm r-LASSO. The limit of the solution as $\lambda \to 0$ is

$$\arg\min_{\tilde{x}} \|\tilde{x}\|_1, \quad \text{s.t. } \tilde{y} = \tilde{A}\tilde{x},$$

which is called the basis pursuit problem or r-BP in this paper. It is straightforward to extend the analyses of LASSO and BP for the real signals to r-LASSO and r-BP.[1] However, this approach ignores the information about the grouping of the real and imaginary parts. In fact, in many applications the real and imaginary components tend to be small or large simultanously.

- c-LASSO -- The more natural extension of the LASSO to the complex setting is the following optimization problem that we term c-LASSO

$$\min \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1,$$

where the complex ℓ_1-norm is defined as $\|x\|_1 = \sum_i |x_i| = \sum_i \sqrt{(x_i^R)^2 + (x_i^I)^2}$ [4,5,15–17]. The limit of the solution as $\lambda \to 0$ is

$$\arg\min_x \|x\|_1, \quad \text{s.t. } y = Ax,$$

which we refer to as c-BP.

An important questions we address in this paper is: can we measure how much the grouping of the real and the imaginary parts improves the performance of c-LASSO? Several papers have considered similar problems [18–35] and have provided guarantees on the performance of this algorithm. However, the results are usually inconclusive because of the loose constants involved in their analysis. This paper aims to address the above questions with an analysis that does not involve any loose constants and therefore provides accurate comparisons.

Motivated by the recent results in the asymptotic analysis of the LASSO [3], [7], we first derive the complex approximate message passing algorithm (CAMP) as a fast and efficient algorithm for solving the c-LASSO problem. We then extend the state evolution (SE) framework introduced in [3] to predict the performance of the CAMP algorithm in the asymptotic setting. Since the CAMP algorithm solves c-LASSO, such predictions are accurate for c-LASSO as well for $N \to \infty$. The analysis carried out in this paper provides new information and insight on the performance of the c-LASSO that was not known before such as the least favorable distribution and the noise sensitivity of c-LASSO and CAMP. A more detailed description of the contributions of this paper is summarized in Section 1.5.

1.3 Notation

Let $|\alpha|$, $\measuredangle\alpha$ and α^* denote the amplitude, phase, and conjugate of $\alpha \in \mathbb{C}$ respectively. Furthermore, for the matrix $A \in \mathbb{C}^{n\times N}$, A^*, A_i, A_{ij} denote the conjugate transpose, i^{th} column and ij^{th} element of matrix A. We are interested in approximating a sparse signal $s_o \in \mathbb{C}^N$ from an undersampled set of noisy linear measurements $y = As_o + w$. $A \in \mathbb{C}^{n\times N}$ has iid random elements (with independent real and imaginary parts) from a given distribution that satisfies $\mathbb{E}A_{ij} = 0$ and $E|A_{ij}|^2 = \frac{1}{n}$, and $w \in \mathbb{C}^N$ is the measurement noise. Throughout the paper, we assume that the noise is iid $CN(0, \sigma^2)$, where CN stands for the complex normal.

We are interested in the asymptotic setting where $\delta = n/N$ and $\rho = k/n$ are fixed, while $N \to \infty$. We further assume that the elements of s_o are iid $s_{o,i} \sim (1-\rho\delta)\delta_0(s_{o,i}) + \rho\delta G(s_{o,i})$, where G is an unknown probability distribution without any point mass at 0.[2] Clearly, the expected number of non-zero elements

[1] The theoretical results on LASSO and BP consider iid Gaussian measurement matrices [13]. However, it is conjectured that the results are universal and hold for a "larger" class of random matrices [7, 14].

[2] This assumption is not necessary and as long as the marginal distribution of s_o converges to a given distribution the statements of this paper will hold. For further information on this refer to [7] and [10].

in the vector s_o is $\rho\delta N$. In this model we are assuming that all the non-zero real and imaginary coefficients are paired. This quantifies the maximum amount of improvement the c-LASSO gains by grouping the real and imaginary parts.

Define $\mathcal{F}_{\epsilon,\gamma}$ as the family of distributions F with $F(0^+) - F(0^-) > 1 - \epsilon$ and $E_F(X^2) \leq \epsilon\gamma^2$. Also, define $\mathcal{F}_\epsilon = \{ F \mid F(0^+) - F(0^-) \leq \epsilon\}$.

1.4 Performance criteria

We compare the c-LASSO with r-LASSO in both the noise-free and noisy measurements cases. For each scenario, we define a specific measure to compare the performance of the two algorithms.

1.4.1 Noise-free measurements

Let $\mathcal{A}_\alpha$ be a sparse recovery algorithm with free parameter α. Given (y, A), $\mathcal{A}_\alpha$ returns an estimate $\hat{x}^{\mathcal{A}_\alpha}$ of s_o. Suppose that in the noise free case, as $N \to \infty$, the performance of $\mathcal{A}_\alpha$ exhibits a sharp phase transition, i.e., for every value of δ, there exists $\rho^{\mathcal{A}_\alpha}(\delta)$, below which $\lim_{N\to\infty} \|\hat{x}^{\mathcal{A}_\alpha} - s_o\|^2/N \to 0$ almost surely, while for $\rho > \rho^{\mathcal{A}_\alpha}(\delta)$, $\mathcal{A}_\alpha$ fails with probability 1. The phase transition has been studied either empirically or theoretically for many sparse recovery algorithms [6, 13, 14, 36–39]. The phase transition curve $\rho^{\mathcal{A}_\alpha}(\delta)$ specifies the fundamental exact recovery limit of algorithm $\mathcal{A}_\alpha$.
The free parameter α often affects the performance of the sparse recovery algorithm [6]. Therefore, optimal tuning of this parameter is essential in practical applications. One approach is to tune the parameter for the highest phase transition [6],[3] i.e.,

$$\rho^{\mathcal{A}}(\delta) = \sup_\alpha \rho^{\mathcal{A}_\alpha}(\delta).$$

In other words, $\rho^{\mathcal{A}}$ is the best performance $\mathcal{A}_\alpha$ provides in the exact sparse signal recovery problem, if we know how to tune the algorithm properly. Based on this framework, we say algorithm $\mathcal{A}$ outperforms $\mathcal{B}$ at a given δ, if and only if $\rho^{\mathcal{A}}(\delta) > \rho^{\mathcal{B}}(\delta)$.

1.4.2 Noisy measurements

In the presence of measurement noise exact recovery is not possible. Therefore, tuning the parameter for the highest phase transition curve does not necessarily provide the optimal performance. In this section, we explain the *optimal noise sensitivity tuning* introduced in [7]. Consider the ℓ_2-norm as a measure for the reconstruction error and assume that $\frac{\|\hat{x}^{\mathcal{A}_\alpha} - s_o\|_2^2}{N} \to \mathrm{MSE}(\rho, \delta, \alpha, \sigma)$ almost surely. Define the noise sensitivity of the algorithm as

$$\mathrm{NS}(\rho, \delta, \alpha) = \sup_{\sigma>0} \sup_G \frac{MSE(\rho, \delta, \alpha, \sigma)}{\sigma^2},$$

where α denotes the tuning parameter of algorithm $\mathcal{A}_\alpha$. If the noise sensitivity is large, then the measurement noise may severely degrade the final reconstruction. Therefore, we tune the parameter α to obtain the lowest noise sensitivity, i.e.,

$$\mathrm{NS}(\rho, \delta) = \inf_\alpha \mathrm{NS}(\rho, \delta, \alpha).$$

Based on this framework, we say algorithm $\mathcal{A}$ outperforms $\mathcal{B}$ at a given δ and ρ, if and only if $NS^{\mathcal{A}}(\delta, \rho) < NS^{\mathcal{B}}(\delta, \rho)$.

[3] In this paper, we are considering algorithms whose phase transitions do not depend on the distribution G of non-zero coefficients. Otherwise, one could use the maximin framework introduced in [6].

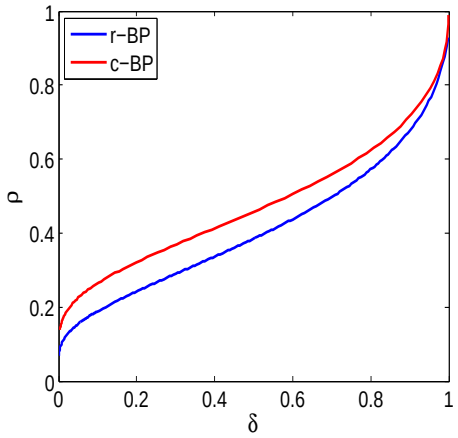

Figure 1: Comparison of the phase transition curve of the r-BP and c-BP. When all the non-zero real and imaginary parts of the signal are grouped, the phase transition of c-BP outperforms that of r-BP.

1.5 Contributions

In this paper, we first develop the *complex approximate message passing* (CAMP) algorithm. This algorithm provides a simple and fast converging iterative method for solving c-LASSO. We extend the *state evolution*, introduced recently as a framework for accurate asymptotic predictions of the AMP performance, to the CAMP. We will then use the connection between CAMP and c-LASSO to provide an accurate asymptotic analysis of the c-LASSO algorithm. We aim to characterize the phase transition curve (noise free measurements) and noise sensitivity (noisy measurements) of these two algorithm when the real and imaginary parts are all paired, i.e., they are both zero or non-zero. Both criteria have been extensively studied for the real signals (and hence for the r-LASSO) [3, 7]. The results of our predictions are summarized in Figures 1, 2, and 3. Figure 1 compares the phase transition curve of c-BP and CAMP with the phase transition curve of r-BP. As we expected c-BP and CAMP outperform r-BP since they exploit the connection between the real and imaginary parts. If $\rho_{SE}(\delta)$ denotes the phase transition curve, we also prove that $\rho_{SE}(\delta) \sim \frac{1}{\log(1/2\delta)}$ as $\delta \to 0$. Comparing this with $\rho^R_{SE}(\delta) \sim \frac{1}{2\log(1/\delta)}$ for the r-LASSO, we conclude that

$$\lim_{\delta\to 0} \frac{\rho_{SE}(\delta)}{\rho^R_{SE}(\delta)} = 2.$$

Figure 2 exhibits the noise sensitivity of c-LASSO and CAMP. We prove in Section 3.3 that, as the sparsity approaches the phase transition curve, the noise sensitivity grows up to infinity. Finally, Figure 3 compares the contour plots of the noise sensitivity of c-LASSO with those of the r-LASSO. For the fixed value noise sensitivity, the level set of the c-LASSO is higher than that of r-LASSO. It is worth noting that the same comparisons are true between CAMP and AMP, as will be clarified in Section 3.4.

1.6 Related work

The state evolution framework used in this paper was first introduced in [3]. Deriving the phase transition and noise sensitivity of the LASSO for real signals and real measurements from SE is due to [7]; see [40] for more comprehensive discussion. Finally, the derivation of the AMP from the full sum-product message

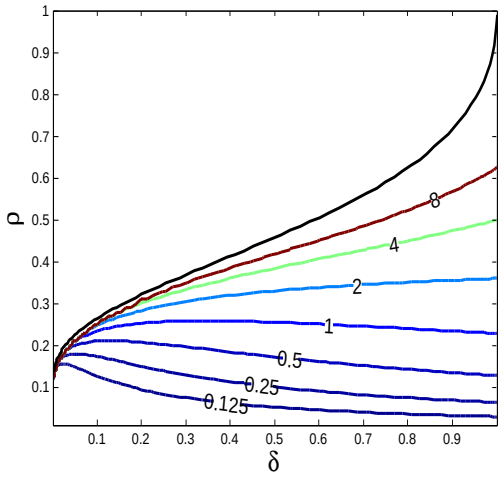

Figure 2: Contour lines of noise sensitivity in the (δ, ρ) plane. The black curve is the phase transition curve at which the noise sensitivity is infinite. The colored lines display the level sets of $\mathrm{NS}(\rho, \delta) = 0.125, 0.25, 0.5, 1, 2, 4, 8$.

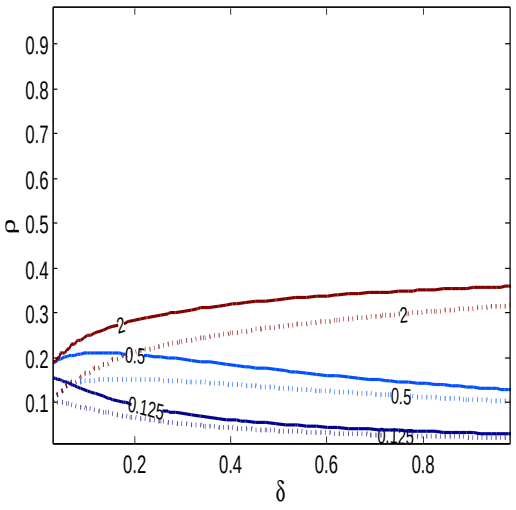

Figure 3: Comparison of the noise sensitivity of r-LASSO with the noise sensitivity of c-LASSO. The colored solid lines present the level sets of the $\mathrm{NS}(\rho, \delta) = 0.125, 0.5, 2$ for the c-LASSO and the colored dotted lines display the same level sets for the r-LASSO.

passing is due to [41]. Our main contribution in this paper is to extend these results to the complex setting. Not only is the analysis of the state evolution more challenging in this setting, but it also provides new insights on the performance of c-LASSO that has not been available before. For instance, the noise sensitivity of c-LASSO is not yet known.

The recovery of sparse complex signals is a special case of group-sparsity or block-sparsity. According to the group sparsity assumption, the non-zero elements of the signal tend to occur in groups or clusters. One of the algorithms used in this context is the group-LASSO [29, 31]. Consider a signal $s_o \in \mathbb{R}^N$. Partition the indices of s_o into m groups $g_1, \ldots, g_m$. The group-LASSO algorithm minimizes the following cost function:

$$\min_x \frac{1}{2}\|y - Ax\|_2^2 + \sum_{i=1}^{m} \lambda_i \|x_{g_i}\|_2, \tag{1}$$

where the λ_i's are regularization parameters. This algorithm has been studied in the literature extensively [18–35]. We briefly review several papers and emphasize on the differences with our work. [32] analyzes the consistency of the group LASSO estimator in the presence of noise. Fixing the signal s_o, it provides conditions under which the group LASSO is consistent as $n \to \infty$. [33, 42] consider a weak notion of consistency that is the exact support recovery. However, [42] proves that in the setting we are interested in, i.e., $k/n = \rho$ and $n/N = \delta$, even the exact support recovery is not possible. Clearly, when the noise is present our goal is neither the exact recovery nor the exact support recovery. Instead, we characterize the mean square error of the reconstruction. This criterion has been considered in [18, 34]. Although the results of [18, 34] show qualitatively the benefit of group sparsity, they do not characterize the difference quantitatively. In fact, loose constants in both the error bound and the number of samples do not permit accurate comparison of the performances. In our analysis, no loose constant is involved, and we provide very accurate characterization of the mean square error.

Group-sparsity and group-LASSO are also of interest in the sparse recovery community. For example, the analysis carried out in [20, 23, 24] are based on "coherence". These results provide sufficient conditions with again loose constants as discussed before. The work of [25–27] addresses this issue by an accurate analysis of the algorithm in the noiseless setting $\sigma = 0$. They provide a very accurate estimate of the phase transition curve for the group-LASSO. However, SE provides a more flexible framework to analyze c-LASSO than the analysis of [27], and it provides more information than just the phase transition curve. For instance, it shows the least favorable distribution of the input and noise sensitivity of c-LASSO.

While writing this paper we were made aware that in an independent work Donoho and Montanari are extending the state evolution framework to the general setting of group sparsity [43]. Their work considers the state evolution framework for the group-LASSO problem and will include the generalization of the analysis provided in this paper to the case where the variables tend to cluster in groups of size m.

Also, both the complex signals and group-sparse signals are special cases of model based compressed sensing (CS) [44]. By introducing more structured models for the signal, [44] proves that the number of measurements needed are proportional to the "complexity" of the model rather than the sparsity level. The results in model-based CS also suffer from loose constants in both the number of measurements and the mean square error bounds.

Finally, from the algorithmic point of view, several papers have considered solving the c-LASSO prob-

lem using first-order methods [4, 15].[4] The deterministic framework that measures the convergence of an algorithm on the most difficult problem instance, that gives the slowest convergence rate, is not an appropriate measure of the convergence rate for the compressed sensing problems [9]. Therefore, [9] considers the average convergence rate for iterative algorithm. In that setting, AMP is the only first order algorithm that provably achieves linear convergence to date. Similarly, the CAMP algorithm, introduced in this paper, provides the first, first-order c-LASSO solver that provides a linear average convergence rate.

1.7 Organization of the paper

We introduce the CAMP algorithm in Section 2. We then explain the state evolution equations to characterize the evolution of the mean square error through the iterations of the CAMP algorithm in Section 3 and we analyze the important properties of the state evolution equations. We then discuss the connection between our calculations and the solution of LASSO in Section 3.4. We confirm our results thru running Monte Carlo simulations in Section 4.

2 Complex Approximate Message Passing

The high computational complexity of interior point methods for solving large convex optimization problems spurred the development of first-order methods for solving the LASSO problem. See [9] and the references therein for a description of some of these algorithms. One of the most successful algorithms for the compressed sensing problems is the AMP algorithm introduced in [3]. In this section, we use the approach introduced in [41] to derive the approximate message passing algorithm for the c-LASSO problem that we term *Complex Approximate Message Passing* (CAMP).

Let $s_1, s_2, \ldots, s_N$ be N random variables with the following distribution:

$$p(s_1, s_2, \ldots, s_N) = \frac{1}{Z(\beta)} \mathrm{e}^{-\beta\lambda\|s\|_1 - \frac{\beta}{2}\|y-As\|_2^2}, \tag{2}$$

where β is a constant and $Z(\beta) = \int_s \mathrm{e}^{-\beta\lambda\|s\|_1 - \frac{\beta}{2}\|y-As\|_2^2} ds$. As $\beta \to \infty$ the mass of this distribution concentrates around the solution of the LASSO. Therefore, one way to find the solution of LASSO is to marginalize this distribution. However, calculating the marginal distribution is an NP-complete problem. The sum-product message passing algorithm provides a successful heuristic for approximating the marginal distribution. As $N \to \infty$ and $\beta \to \infty$ the iterations of the sum-product message passing algorithm are simplified to [41]

$$\begin{aligned} x_{i\to a}^{t+1} &= \eta\Big(\sum_{b\neq a} A_{bi}^* z_{b\to i}^t; \tau_t\Big), \\ z_{a\to i}^t &= y_a - \sum_{j\neq i} A_{aj} x_{j\to a}^t, \end{aligned} \tag{3}$$

where $\eta(u+iv;\lambda) = \Big(u+iv - \frac{\lambda(u+iv)}{\sqrt{u^2+v^2}}\Big)_+ \mathbb{1}_{\{u^2+v^2>\lambda^2\}}$ is the proximity operator of the complex ℓ_1-norm and is called complex soft thresholding. See Appendix A for further information regarding this function. τ_t is the threshold parameter at time t. The choice of this parameter will be discussed in Section 3.1. The

[4] First-order methods are iterative algorithms that use either the gradient or the subgradient of the function at the previous iterations to update their estimates.

per iteration computational complexity of this algorithm is high since $2nN$ messages $x^t_{i\to a}$ and $z^t_{a\to i}$ are updated. Therefore, following [41] we assume that

$$\begin{aligned} x^t_{i\to a} &= x^t_i + \delta x^t_{i\to a} + O(1/N), \\ z^t_{a\to i} &= z^t_a + \delta z^t_{i\to a} + O(1/N), \end{aligned} \tag{4}$$

where $\delta x^t_{i\to a}, \delta z^t_{i\to a} = O(\frac{1}{\sqrt{N}})$. Here, $O(\cdot)$ errors are uniform in the choice of the edges $i \to a$ and $a \to i$. For further discussion of this assumption and its validation see [41]. Let η^I and η^R be the imaginary and real parts of the complex soft thresholding function. Furthermore, define $\frac{\partial \eta^R}{\partial x}$, and $\frac{\partial \eta^R}{\partial y}$ as the partial derivatives of η^R with respect to the real and imaginary parts of the input respectively. $\frac{\partial \eta^I}{\partial x}$, and $\frac{\partial \eta^I}{\partial y}$ are defined similarly. We then have

Proposition 2.1. *Suppose that* (4) *holds for every iteration of the message passing algorithm specified in* (3)*. Then x^t_i and z^t_a satisfy the following equations:*

$$\begin{aligned} x^{t+1}_i &= \eta\Big(x^t_i + \sum_b A^*_{bi} z^t_b; \tau_t\Big), \\ z^t_a &= y_a - \sum_j A_{aj} x^t_j - \sum_j A_{aj} \left(\frac{\partial \eta^R}{\partial x}(x^{t-1}_j + \sum_b A^*_{b,j} z^{t-1}_b)\right) \mathcal{R}(A^*_{a,j} z^t_a) \\ &\quad - \sum_j A_{aj} \left(\frac{\partial \eta^R}{\partial y}(x^{t-1}_j + \sum_b A^*_{b,j} z^{t-1}_b)\right) \mathcal{I}(A^*_{a,j} z^t_a) \\ &\quad - i\sum_j A_{aj} \left(\frac{\partial \eta^I}{\partial x}(x^{t-1}_j + \sum_b A^*_{b,j} z^{t-1}_b)\right) \mathcal{R}(A^*_{a,j} z^t_a) \\ &\quad - i\sum_j A_{aj} \left(\frac{\partial \eta^I}{\partial y}(x^{t-1}_j + \sum_b A^*_{b,j} z^{t-1}_b)\right) \mathcal{I}(A^*_{a,j} z^t_a). \end{aligned} \tag{5}$$

where $\mathcal{R}$ and $\mathcal{I}$ operators return the real and imaginary parts of a complex number respectively.

See Appendix B for the proof. According to Proposition 2.1 and (4), for large values of N, the messages $x^t_{i\to a}$ and $z^t_{a\to i}$ are close to x^t_i and z^t_a in (5). Therefore, we define the CAMP algorithm as the iterative method that starts from $x^0 = 0$ and $z^0 = y$ and uses the iterations specified in (5). It is important to note that Proposition 2.1 does not provide any information on either the performance of the CAMP algorithm or the connection between CAMP and c-LASSO, since the message passing is a heuristic algorithm and does not necessarily converge to the correct marginal distribution of (2).

3 Formal analysis of CAMP and c-LASSO

In this section, we explain the state evolution framework as a framework that predicts the performance of the CAMP and c-LASSO in the asymptotic settings. We then use this framework to analyze the phase transition and noise sensitivity of the c-LASSO and CAMP.

3.1 State evolution

We now conduct an asymptotic analysis of the CAMP algorithm. As we confirm in Section 3.4, the asymptotic performance of the algorithm is tracked through a few variables, called the state variables. The

state of the algorithm is the 5-tuple $\mathbf{s} = (m; \delta, \rho, \sigma, G)$, where G corresponds to the distribution of the non-zero elements of the sparse vector s_o, σ is the standard deviation of the measurement noise, and m is the asymptotic normalized mean square error. The threshold parameter (threshold policy) in its most general form could be a function of the state of the algorithm $\tau(\mathbf{s})$. Define $\text{npi}(m; \sigma, \delta) \triangleq \sigma^2 + \frac{m}{\delta}$. The MSE map is defined as

$$\Psi(\mathbf{s}, \tau(\mathbf{s})) \triangleq \mathbb{E}|\eta(X + \sqrt{\text{npi}(m, \sigma, \delta)}Z_1 + i\sqrt{\text{npi}(m, \sigma, \delta)}Z_2; \tau(\mathbf{s})) - X|^2,$$

where $Z_1, Z_2 \sim N(0, 1/2)$, and $X \sim (1 - \rho\delta)\delta_0(x) + \rho\delta G(x)$ are independent random variables. Note that G is a probability distribution on $\mathbb{C}$. In the rest of this paper, we consider the thresholding policy $\tau(\mathbf{s}) = \tau\sqrt{\text{npi}(m, \sigma, \delta)}$, where the constant τ is yet to be tuned according to the schemes introduced in Sections 1.4.1 and 1.4.2. When we use this thresholding policy we may equivalently write $\Psi(\mathbf{s}, \tau(\mathbf{s}))$ as $\Psi(\mathbf{s}, \tau)$. This thresholding policy is the same as the thresholding policy introduced in [3, 7]. When the parameters $\delta, \rho, \sigma, \tau$ and G are clear from the context, we denote the MSE map by $\Psi(m)$. The state evolution is the evolution of m by the rule

$$m_{t+1} = \Psi(m_t).$$

As will be described in Section 3.4, this equation tracks the normalized MSE of the CAMP algorithm in the asymptotic setting $N \to \infty$. In other words, if m_t is the mean square error of the CAMP algorithm at iteration t, m_{t+1} will be the MSE of CAMP at iteration $t + 1$.

Definition 3.1. *Let Ψ be almost everywhere differentiable. m^* is called a fixed point of Ψ if and only if $\Psi(m^*) = m^*$. Furthermore, a fixed point is called stable if $\frac{d\Psi(m)}{dm}\Big|_{m=m^*} < 1$, and unstable if $\frac{d\Psi(m)}{dm}\Big|_{m=m^*} > 1$.*

It is clear that if m^* is the unique stable fixed point of the Ψ function, then $m_t \to m^*$ as $t \to \infty$. Also, if all the fixed points of Ψ are unstable, then $m_t \to \infty$ as $t \to \infty$. Let $\mu = |X|$ and $\theta = \measuredangle X$. Also, let $G(\mu, \theta)$ denote the probability density function of X and $G(\mu) = \int G(\mu, \theta)d\theta$.

Lemma 3.2. *The MSE map does not depend on the phase distribution of the input signal, i.e.,*

$$\Psi(m, \delta, \rho, \sigma, G(\mu, \theta), \tau) = \Psi(m, \delta, \rho, \sigma, G(\mu), \tau).$$

See Appendix C for the proof.

3.2 Noise-free signal recovery

Suppose that the measurements are noise-free, i.e., $\sigma = 0$. Fix all the state variables except for m, and ρ. The evolution of m, discriminates the following two regions for ρ:

Region I : The values of ρ for which $\Psi(m) < m$ for every $m > 0$;

Region II: The complement of Region I.

Since 0 is necessarily a fixed point of the Ψ function, in Region I $m_t \to 0$ as $t \to \infty$. The following lemma shows that in the second region $m = 0$ is an unstable fixed point and therefore starting from $m_0 \neq 0$, $m_t \nrightarrow 0$.

Lemma 3.3. *Let $\sigma = 0$. If ρ is in Region II, then Ψ has an unstable fixed point at zero.*

Proof. We prove in Lemma D.1 that $\Psi(m)$ is a concave function of m. Therefore, ρ is in Region II if and only if $\frac{d\Psi(m)}{dm}\Big|_{m=0} > 1$. This in turn indicates that 0 is an unstable fixed point. □

It is also easy to confirm that Region I is of the form $[0, \rho_{SE}(\delta, G, \tau))$. As we will see in Section 3.4, $\rho_{SE}(\delta, G, \tau)$ determines the phase transition curve of the CAMP algorithm. According to Lemma 3.2, the MSE map does not depend on the phase distribution of the non-zero elements. However, it depends on $G(\mu)$. The following proposition shows that in fact ρ_{SE} is independent of G even though the Ψ function does depend on $G(\mu)$.

Proposition 3.4. *$\rho_{SE}(\delta, G, \tau)$ is independent of the distribution G.*

Proof. According to Appendix D Ψ is concave. Therefore, it has a stable fixed point at zero if and only if its derivative at zero is less than 1. It is also straightforward (from Appendix D) to show that

$$\frac{d\Psi}{dm}\Big|_{m=0} = \frac{\rho\delta(1+\tau^2)}{\delta} + \frac{1-\rho\delta}{\delta}\mathbb{E}|\eta(Z_1 + iZ_2; \tau)|^2.$$

Setting this derivative to 1, it is clear that the phase transition value of ρ is independent of G. □

According to Proposition 3.4 the only parameters that affect ρ_{SE} are δ and the free parameter τ. As mentioned in Section 1.4.1, one approach in tuning τ is to set it such that the algorithm achieves its highest phase transition, i.e.,

$$\rho_{SE}(\delta) = \sup_{\tau} \rho_{SE}(\delta; \tau).$$

Using the state evolution we can calculate the optimal value of τ and $\rho_{SE}(\delta)$.

Theorem 3.5. *$\rho_{SE}(\delta)$ and δ satisfy the following implicit relations:*

$$\begin{aligned}
\rho_{SE}(\delta) &= \frac{\int_{\omega\geq\tau} \omega(\tau-\omega)\mathrm{e}^{-\omega^2} d\omega}{(1+\tau^2)\int_{\omega>\tau} \omega(\tau-\omega)\mathrm{e}^{-\omega^2} d\omega - \tau\int_{\omega>\tau} \omega(\omega-\tau)^2\mathrm{e}^{-\omega^2}}, \\
\delta &= \frac{4(1+\tau^2)\int_{\omega>\tau} \omega(\tau-\omega)\mathrm{e}^{-\omega^2} d\omega - 4\tau\int_{\omega>\tau} \omega(\omega-\tau)^2\mathrm{e}^{-\omega^2} d\omega}{-2\tau + 4\int_{\omega>\tau} \omega(\tau-\omega)\mathrm{e}^{-\omega^2} d\omega},
\end{aligned}$$

for $\tau \in [0, \infty)$.

See Appendix E for the proof. Figure 1 displays this phase transition curve. The implicit formulation above allows us to calculate the asymptotic performance of the phase transition as $\delta \to 0$.

Theorem 3.6. *$\rho_{SE}(\delta)$ follows the the asymptotic behavior*

$$\rho_{SE}(\delta) \sim \frac{1}{\log\left(\frac{1}{2\delta}\right)}, \quad \text{as} \quad \delta \to 0.$$

See Appendix F for the proof. As mentioned before, this theorem shows that as $\delta \to 0$ the phase transition of c-LASSO and CAMP is two time that of the r-LASSO. This is the benefit of grouping the real and imaginary parts in the noise-free case. We discuss optimal strategies for noisy cases in Section 3.3.

3.3 Noise sensitivity

We first discuss the risk of the complex soft thresholding function. The properties of this risk play an important role in the discussion of the noise sensitivity of state evolution in Section 3.3.2.

3.3.1 Risk of soft thresholding

Define the risk of the soft thresholding function as

$$r(\mu, \tau) = \mathbb{E}|\eta(\mu e^{j\phi} + Z_1 + iZ_2; \tau) - X|^2,$$

where the expected value is with respect to the two independent random variables $Z_1, Z_2 \sim N(0, 1/2)$. It is important to note that according to Lemma 3.2 the risk function is independent of ϕ. The following lemma characterizes two important properties of this risk function:

Lemma 3.7. *$r(\mu, \tau)$ is an increasing function of μ and a concave function in terms of μ^2.*

See Appendix G for the proof of this lemma. We define the minimax risk of the soft thresholding function as

$$M^\flat(\epsilon) = \inf_{\tau>0} \sup_{\nu \in F_\epsilon} \mathbb{E}|\eta(X + Z_1 + iZ_2; \tau) - X|^2,$$

where ν is the distribution of X, Z_1 and Z_2.

Proposition 3.8. *The minimax risk of the soft thresholding function satisfies,*

$$M^\flat(\epsilon) = \inf_{\tau} \; 2(1-\epsilon) \int_{w=\tau}^{\infty} w(w-\tau)^2 e^{-w^2} dw + \epsilon(1+\tau^2).$$

See Appendix H for the proof. We use this minimax risk to derive the noise sensitivity of the state evolution in the next section.

3.3.2 Noise sensitivity of state evolution

As mentioned in Section 3.1, in the presence of measurement noise the state evolution is given by

$$\begin{aligned} m_{t+1} &= \Psi_\sigma(m_t), \\ \Psi_\sigma(m_t) &= \mathbb{E}|\eta(X + \sqrt{\text{npi}} Z_1 + i\sqrt{\text{npi}} Z_2; \tau\sqrt{\text{npi}}) - X|^2, \end{aligned}$$

where $\text{npi} = \sigma^2 + \frac{m_t}{\delta}$.

Lemma 3.9. *$\Psi_\sigma(m)$ has a unique stable fixed point to which the sequence of $\{m_t\}$ converges.*

We call this fixed point $\text{fMSE}(\sigma^2, \delta, \rho, G, \tau)$. According to Section 1.4.2, we define the minimax noise sensitivity as

$$NS^{SE}(\delta, \rho) = \min_{\tau} \sup_{\sigma>0} \sup_{\nu \in \mathcal{F}_\epsilon} \text{fMSE}(\sigma^2, \delta, \rho, G, \tau)/\sigma^2.$$

The noise sensitivity of the state evolution can be easily evaluated from $M^\flat(\epsilon)$. The following theorem characterizes this relation.

Theorem 3.10. *Let $\rho_{MSE}(\delta)$ be the value of ρ satisfying $M^\flat(\rho\delta) = \delta$. Then, for $\rho < \rho_{MSE}$ we have*

$$NS^{SE}(\delta, \rho) = \frac{M^\flat(\delta\rho)}{1 - M^\flat(\delta\rho)/\delta},$$

and for $\rho > \rho_{MSE}(\delta)$, $\text{NS}^{SE}(\delta, \rho) = \infty$.

The proof of this theorem follows exactly the same lines as the proof of Proposition 3.1 in [7], and therefore we skip it for the sake of brevity. The contour lines of this noise sensitivity function are displayed in Figure 2.

Lemma 3.11. *For every $\delta \in [0,1]$ we have,*

$$\rho_{MSE}(\delta) = \rho_{SE}(\delta).$$

Proof. The proof is a simple comparison of the formulas. We first know that ρ_{MSE} is derived from the following equation

$$\min_{\tau} 2(1-\rho\delta)\int_{\omega>\tau}\omega(\omega-\tau)^2 \mathrm{e}^{\omega^2} d\omega + \rho\delta(1+\tau^2) = \delta.$$

On the other hand, since $\Psi(m)$ is a concave function of m, $\rho_{SE}(\delta,\tau)$ is derived from $\frac{d\Psi(m)}{dm}\Big|_{m=0} = 1$. This derivative is equal to

$$\frac{d\Psi(m)}{dm}\Big|_{m=0} = \frac{2(1-\rho\delta)}{\delta}\int_{\omega>\tau}\omega(\omega-\tau)^2\mathrm{e}^{\omega^2}d\omega + \frac{\rho\delta}{\delta}(1+\tau^2).$$

Also, $\rho_{SE}(\delta) = \sup_\tau \rho_{SE}(\tau,\delta)$. However, in order to obtain the highest ρ we should minimize the above expression over τ. Therefore, both $\rho_{SE}(\delta)$ and $\rho_{MSE}(\delta)$ satisfy the same equations and are exactly the same. □

3.4 Connection between the state evolution, CAMP, and c-LASSO

There is a strong connection between the state evolution framework and the CAMP algorithm and c-LASSO. Recently, [10] proved that the state evolution predicts the asymptotic performance of the AMP algorithm when the measurement matrix is iid Gaussian. The result also holds for complex Gaussian matrices and complex input vectors, and the proof is essentially the same as the real case. As in [3], we conjecture that the SE predictions are correct for a "wide" class of random matrices. We show evidence of this claim in Section 4. Here, for the sake of completeness, we quote the result of [10] in the complex setting. Let $\gamma : \mathbb{C}^2 \to \mathbb{R}$ be a pseudo Lipschitz function,[5] and s_o and x^t denote the original vector and the estimates of CAMP at time t respectively. Suppose that the empirical distribution of s_o converges to distribution F_s. We then have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\gamma(x_i^t, s_{o,i}) = \mathbb{E}\gamma\left(\eta\left(X+\sqrt{\mathrm{npi}^t}Z_1 + i\sqrt{\mathrm{npi}^t}Z_2; \tau\sqrt{\mathrm{npi}^t}\right), X\right) \tag{6}$$

almost surely, where $Z_1 + iZ_2 \sim CN(0,1)$ and $X \sim F_s$ are independent complex random variables. It is also simple to extend the result of [7] and [45] on the connection of message passing algorithms and LASSO to the complex setting. For a given value of τ suppose that the fixed point of the state evolution is denoted by m^*. Define $\lambda(\tau)$ as

$$\lambda(\tau) \triangleq \tau\sqrt{m^*}\left(1-\frac{1}{2\delta}\mathbb{E}\Big(\partial_1\eta^R\Big(X+\sqrt{m^*}Z_1+i\sqrt{m^*}Z_2;\tau\sqrt{m^*}\Big)+\partial_2\eta^I\left(X+\sqrt{m^*}Z_1+i\sqrt{m^*}Z_2;\tau\sqrt{m^*}\right)\Big)\right), \tag{7}$$

and suppose that $\hat{x}(\lambda(\tau))$ is the solution of c-LASSO with the regularization parameter $\lambda(\tau)$. We then have

$$\lim_{t\to\infty}\lim_{N\to\infty}\frac{1}{N}\|x^t - \hat{x}(\lambda(\alpha))\|_2^2 = 0$$

almost surely. For more information on the connections between these algorithms, see [7] and [45].

[5] $\gamma : \mathbb{C}^2 \to \mathbb{R}$ is pseudo-Lipschitz if and only if $|\psi(x)-\psi(y)| \le L(1+\|x\|_2+\|y\|_2)\|x-y\|_2$.

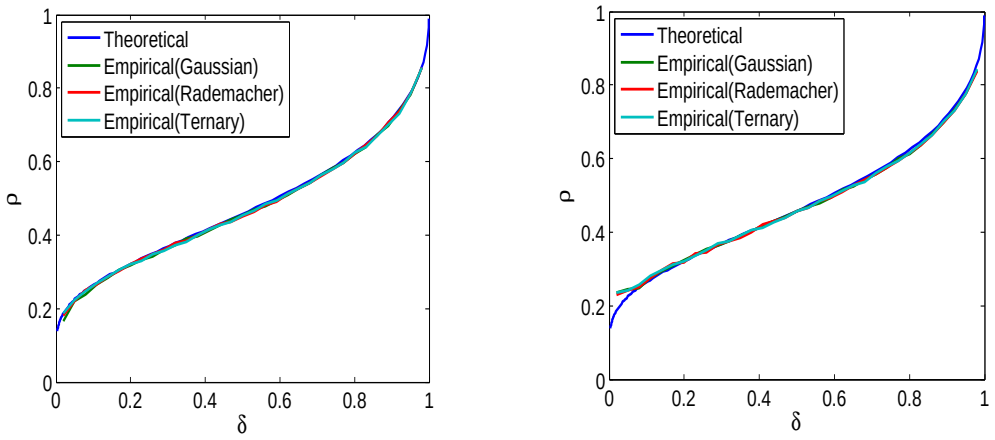

Figure 4: Comparison of $\rho_{SE}(\delta)$ with the empirical phase transition of c-LASSO [46] (left) and CAMP (right). There is a close match between the theoretical prediction and the empirical results from Monte Carlo simulations.

4 Simulations

As explained in Section 3.4, our theoretical results show that if the elements of the matrix are iid Gaussian then state evolution predicts the performance of the CAMP and c-LASSO algorithms accurately. However, in this section we will show evidence that suggests the theoretical framework is applicable to a wider class of measurement matrices. We then investigate the dependence of the empirical phase transition on the input distribution for medium problem sizes.

4.1 Measurement matrix simulations

We investigate the effect of the measurement matrix distribution on the performance of CAMP and c-LASSO in two different cases. First, we consider the case where the measurements are noise-free. We postpone a discussion of measurement noise to Section 4.1.2.

4.1.1 Noise-free measurements

Suppose that the measurements are noise-free. Our goal is to empirically measure the phase transition curves of the c-LASSO and CAMP on the measurement matrices provided in Table 1. To characterize the phase transition of an algorithm, we do the following:

- We consider 33 equispaced values of δ between 0 and 1.
- For each value of δ, we calculate $\rho_{SE}(\delta)$ from the theoretical framework and then consider 41 equispaced values of ρ in $[\rho_{SE}(\delta) - 0.2, \rho_{SE}(\delta) + 0.2]$.
- We fix $N = 1000$, and for any value of ρ and δ, we calculate $n = \lfloor \delta N \rfloor$ and $k = \lfloor \rho \delta N \rfloor$.
- We draw $M = 20$ independent random matrices from the given distribution and for each matrix we construct a random input vector s_o with a given distribution. We then form $y = As_o$ and recover s_o from y, A by either c-BP or optimally tuned CAMP to obtain $\hat{x}$. The matrix distributions and coefficient distributions we consider in our simulations are specified in Tables 1 and 2, respectively.

- For each δ, ρ and Monte Carlo sample j we define a success variable $S_{\delta,\rho,j} = \mathbb{I}(\frac{\|\hat{x}-x\|}{\|x\|} < \text{tol})$ and using these variable we calculate the success probability $\hat{p}^S_{\delta,\rho} = \frac{1}{M}\sum_j S_{\delta,\rho,j}$. This provides an empirical estimate of the probability of correct recovery. The value of tol in our case is set to 10^{-4}.

- For a fixed value of δ, we fit a logistic regression function to $\hat{p}^S(\delta,\rho)$ to obtain $p^S_\delta(\rho)$. Then we find the value of $\hat{\rho}_\delta$ for which $p^S_\delta(\rho) = 0.5$.

See [6] for a more detailed discussion of this approach. For the c-LASSO algorithm we are reproducing the experiments of [46] and, therefore, we are using one-L1 algorithm [46]. Although the experiments show that in the regions of the phase diagram where the noise sensitivity is less than 2, even 100 iterations of CAMP are enough for the convergence, since our goal is to measure the phase transition, we consider 3000 iterations.

Figure 4 compares the phase transition of c-LASSO and CAMP on the ensembles specified in Table 1 with the theoretical prediction of this paper. In this simulation the coefficient ensemble is UP. Clearly, the empirical and theoretical phase transitions of the algorithms are very close.

Table 1: Ensembles considered for the measurement matrix A in the matrix universality ensemble experiments.

Name	Specification
Gaussian	iid elements with standard normal real and imaginary parts
Rademacher	iid elements with real and imaginary parts distributed according to $\frac{1}{2}\delta_{-\sqrt{\frac{1}{2n}}}(x) + \frac{1}{2}\delta_{\sqrt{\frac{1}{2n}}}(x)$
Ternary	iid elements real and imaginary parts distributed according to $\frac{1}{3}\delta_{-\sqrt{\frac{3}{2n}}}(x) + \frac{1}{3}\delta_0(x) + \frac{1}{3}\delta_{\sqrt{\frac{3}{2n}}}(x)$

4.1.2 Noisy measurements

In this section we aim to show that, even in the presence of noise, the matrix ensembles defined in Table 1 perform similarly. Here is the setup for our experiment:

- We set $\delta = 0.25$, $\rho = 0.1$, and $N = 1000$.

- We choose 50 different values of σ in the range [0.001, .1].

- We choose $n \times N$ measurement matrix A from one of the ensembles specified in Table 1.

- We draw k iid elements from CAUP ensemble for the $k = \lfloor \rho n \rfloor$ non-zero elements of the input s_o.

- We form the measurement vector $y = As_o + \sigma w$ where w is the noise vector with iid elements from $CN(0,1)$.

- For CAMP we set $\tau = 2$. For c-LASSO we use (7) to derive the corresponding values of λ for $\tau = 2$ in CAMP.

- We calculate the MSE $\|\hat{x} - s_o\|_2^2/N$ for each matrix ensemble and compare the results.

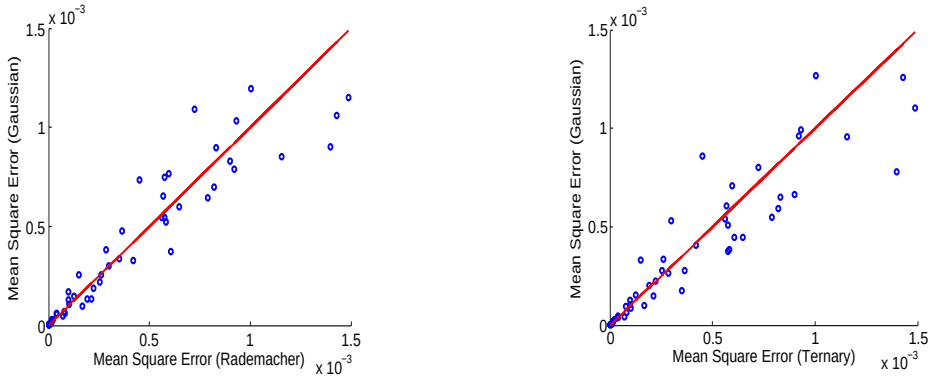

Figure 5: Comparison of the means square error of c-LASSO for Gaussian and Rademacher matrix ensembles (left), and Gaussian and Ternary ensemble (right). The concentration of points around the $y = x$ confirms the universality hypothesis. The norms of residuals are equal to 5.9×10^{-4} and 6×10^{-4} for the left and right figures, respectively. Comparison of this figure with Figure 7 also confirms that as N grows the points become more concentrated about $y = x$ line.

Figures 5 and 6 summarize our results. The concentration of the points along the $y = x$ line indicates that the above matrix ensembles are performing similarly. In order to provide a stronger evidence, we run the above experiment with $N = 4000$. The result of this experiment is exhibited in Figures 7 and 8. It is clear from these figures that the MSE is now more concentrated around the $y = x$ line. Results obtained for other values of the parameters confirm similar behaviors.

4.2 Coefficient ensemble simulations

According to Proposition 3.4, $\rho_{SE}(\delta, \tau)$ is independent of the distribution G of non-zero coefficients of s_0. We test the accuracy of this result on medium problem sizes. We fix δ to 0.1 and we calculate $\hat{p}^S_{\delta,\rho}$ for 60 equispaced values of ρ between 0.1 and 0.5. For each algorithm and each value of ρ we run 100 Monte Carlo trials and calculate the success rate for the Gaussian matrix and the coefficient ensembles specified in Table 2. Figure 9 summarizes our result. Simulations at other values of δ result in very similar behavior. These results are consistent with Theorem 3.4.

Table 2: Coefficient ensembles considered in coefficient ensemble experiments.

Name	Specification
UP	iid elements with amplitude 1 and uniform phase
ZP	iid elements with amplitude 1 and phase zero
GA	iid elements with standard normal real and imaginary parts
UF	iid elements with $U[0, 1]$ real and imaginary parts

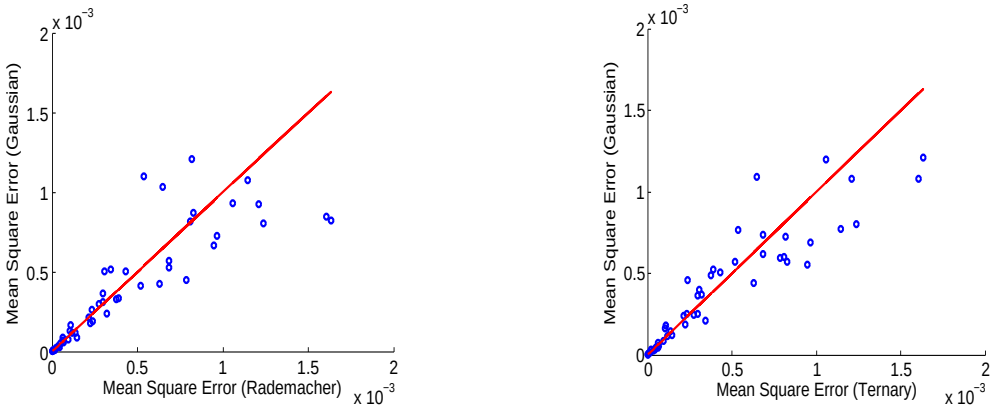

Figure 6: Comparison of the MSE of c-LASSO for Gaussian and Rademacher matrix ensembles (left), and Gaussian and Ternary ensemble (right). The concentration of points around the y = x confirms the universality hypothesis. The norms of residuals are equal to 9.1×10^{-4} and 9.4×10^{-4} for the left and right figures, respectively. Comparison of this figure with Figure 8 also confirms that as N grows the points become more concentrated about $y = x$ line.

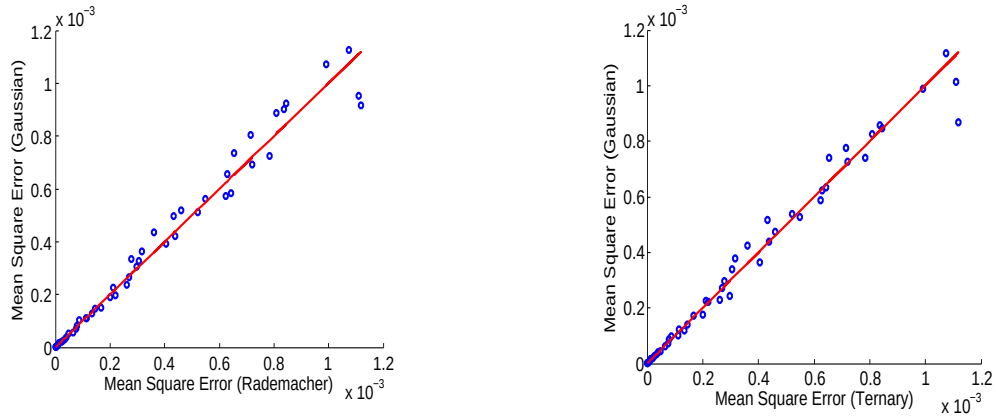

Figure 7: Comparison of the MSE of c-LASSO for Gaussian and Rademacher matrix ensembles (left), and Gaussian and Ternary ensemble (right). The norms of residuals are 2.8×10^{-4} and 2.3×10^{-4}. Compare with Figure 5.

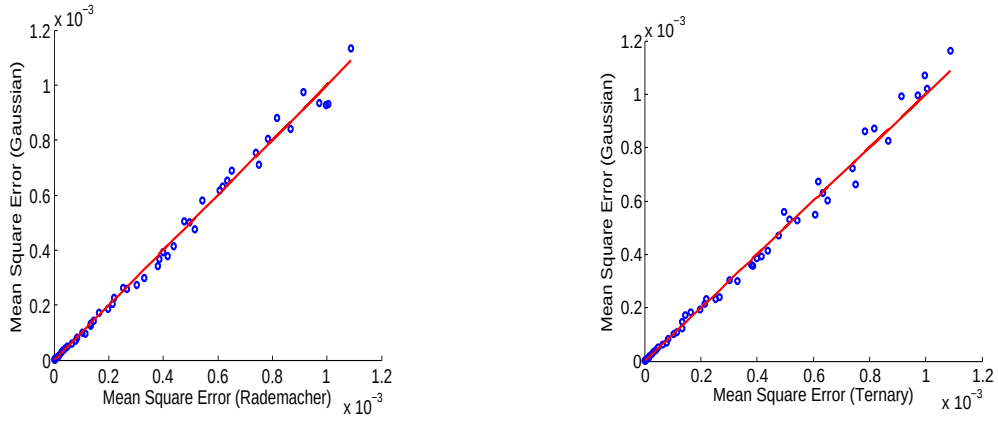

Figure 8: Comparison of the MSE of c-LASSO for Gaussian and Rademacher matrix ensembles (left), and Gaussian and Ternary ensemble (right). The norms of residuals are 2×10^{-4} and 1.8×10^{-4}, respectively. Compare with Figure 6.

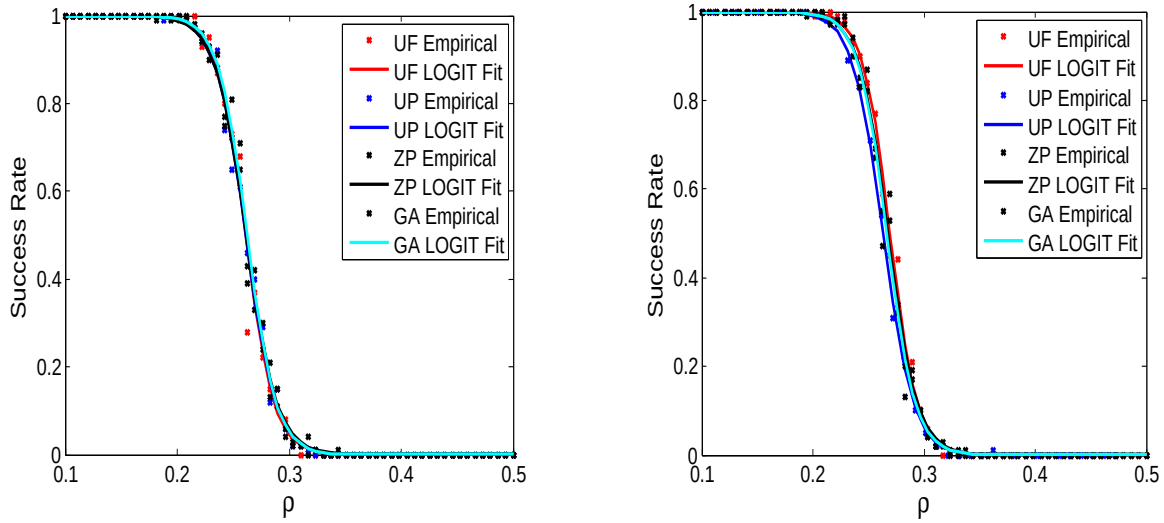

Figure 9: Comparison of the phase transition of c-LASSO (left) and CAMP (right) for different coefficient ensembles specified in Table 2. $\delta = 0.1$ in this figure. These figures are in agreement with the coefficient universality hypothesis.

5 Conclusions

The problem of recovering a complex sparse signal from an undersampled set of complex measurements is considered in this paper. We accurately analyzed the asymptotic performance of c-LASSO and CAMP algorithms. Using the state evolution framework, we proved simple expressions for the noise sensitivity and phase transition of these two algorithms. The results presented here show that substantial improvements can be achieved when the real and imaginary parts are considered jointly in the algorithm. For instance as $\delta \to 0$ we showed that the phase transition of CAMP and c-BP is two times higher than the phase transition of r-LASSO.

6 Acknowledgements

The authors would like to thank David L. Donoho and Andrea Montanari for their encouragement and their valuable suggestions on an early draft.

A Proximity operator

For a given convex function $f : \mathbb{C}^n \to \mathbb{R}$ the proximity operator at point x is defined as

$$\text{Prox}_f(x) = \arg\min_{y\in\mathbb{C}^n} \frac{1}{2}\|y-x\|_2^2 + f(y). \tag{8}$$

The proximity operators play an important role in optimization theory. For further information refer to [47] or Chapter 7 of [40] . The following lemma characterizes the proximity operator for the complex ℓ_1-norm. This proximity operator has been used in several other papers [4, 15–17, 46].

Lemma A.1. *Let f denote the complex ℓ_1-norm function, i.e., $f(x) = \sum_i \sqrt{(x_i^R)^2 + (x_i^I)^2}$. Then the proximity operator is given by*

$$\text{Prox}_{\tau f}(x) = \eta(x;\tau),$$

where $\eta(u+iv;\lambda) = \left(u + iv - \frac{\lambda(u+iv)}{\sqrt{u^2+v^2}}\right)_+ \mathbb{1}_{\{u^2+v^2>\lambda^2\}}$ is applied component-wise to the vector x.

Proof. Since (8) can be decoupled into the elements of the vectors, we solve the optimization for a single component. In other words, we solve the optimization in (8) for $x, y \in \mathbb{C}$. Suppose that the optimal y_* satisfies $(y_*^R)^2 + (y_*^I)^2 > 0$. Then the function $\sqrt{(y^R)^2 + (y^I)^2}$ is differentiable and the optimal solution satisfies

$$\begin{aligned} x^R - y_*^R &= \frac{\lambda y_*^R}{\sqrt{(y_*^R)^2 + (y_*^I)^2}}, \\ x^I - y_*^I &= \frac{\lambda y_*^I}{\sqrt{(y_*^R)^2 + (y_*^I)^2}}. \end{aligned} \tag{9}$$

Combining the two equations in 9 we obtain $y_*^R x^I = x^R y_*^I$. Replacing this in (9) we have $y_*^R = x^R - \frac{\lambda|x^R|}{\sqrt{(x^R)^2+(x^I)^2}}$ and $y_*^I = x^I - \frac{\lambda|x^I|}{\sqrt{(x^R)^2+(x^I)^2}}$. It is clear that if $\sqrt{(x^R)^2 + (x^I)^2} < \lambda$ the signs of y_*^R and x^R will be opposite which is in contradiction with (9). Therefore, if $\sqrt{(x^R)^2 + (x^I)^2} < \lambda$, both y_*^R and y_*^I are zero. It can be easily checked that $(0, 0)$ satisfies the subgradient optimality condition. □

B Proof of Proposition 2.1

Let

$$\begin{aligned}\eta^R(x+iy) &\triangleq \mathcal{R}(\eta(x+iy;\lambda)),\\ \eta^I(x+iy) &\triangleq \mathcal{I}(\eta(x+iy;\lambda)),\end{aligned}$$

denote the real and imaginary parts of the complex soft thresholding function. Define

$$\begin{aligned}\partial_1\eta^R &= \frac{\partial\eta^R(x+iy)}{\partial x}, \qquad \partial_2\eta^R = \frac{\partial\eta^R(x+iy)}{\partial y},\\ \partial_1\eta^I &= \frac{\partial\eta^I(x+iy)}{\partial x}, \qquad \partial_2\eta^I = \frac{\partial\eta^I(x+iy)}{\partial y}.\end{aligned}$$

We first simplify the expression for $z^t_{a\to i}$.

$$z^t_{a\to i} = \underbrace{y_a - \sum_{j\in[N]} A_{aj}x^t_j - \sum_{j\in[N]} A_{aj}\delta x^t_{j\to a}}_{z^t_a} + \underbrace{A_{ai}x^t_i}_{\delta z^t_{a\to i}}.$$

Also,

$$\begin{aligned}x^{t+1}_{i\to a} &= \eta(\sum_{b\in[n]} A^*_{bi}z^t_b + \sum_{b\in[n]} A^*_{bi}\delta z^t_{b\to i} - A^*_{ai}z^t_a;\tau_t) + O(1/N)\\ &= \underbrace{\eta(\sum_{b\in[n]} A^*_{bi}z^t_b + \sum_{b\in[n]} A^*_{bi}\delta z^t_{b\to i};\tau_t)}_{x^t_i}\\ &\quad - \mathcal{R}(A^*_{ai}z^t_a)\partial_1\eta^R(\sum_{b\in[n]} A^*_{bi}z^t_b + \sum_{b\in[n]} A^*_{bi}\delta z^t_{b\to i})\\ &\quad - \mathcal{I}(A^*_{ai}z^t_a)\partial_2\eta^R(\sum_{b\in[n]} A^*_{bi}z^t_b + \sum_{b\in[n]} A^*_{bi}\delta z^t_{b\to i})\\ &\quad - \mathcal{R}(A^*_{ai}z^t_a)\partial_1\eta^I(\sum_{b\in[n]} A^*_{bi}z^t_b + \sum_{b\in[n]} A^*_{bi}\delta z^t_{b\to i})\\ &\quad - \mathcal{I}(A^*_{ai}z^t_a)\partial_2\eta^I(\sum_{b\in[n]} A^*_{bi}z^t_b + \sum_{b\in[n]} A^*_{bi}\delta z^t_{b\to i}) + O(1/N). \qquad (10)\end{aligned}$$

$\sum_b A^*_{bi}\delta z^t_{b\to i} = x^t_i$, since $\delta z_{b\to i} = A_{bi}x^t_i$, and the columns of the matrix are assumed to be normalized. It is also clear that

$$\begin{aligned}\delta x^t_{i\to a} &= - \mathcal{R}(A^*_{ai}z^t_a)\partial_1\eta^R(\sum_{b\in[n]} A^*_{bi}z^t_b + \sum_{b\in[n]} A^*_{bi}\delta z^t_{b\to i})\\ &\quad - \mathcal{I}(A^*_{ai}z^t_a)\partial_2\eta^R(\sum_{b\in[n]} A^*_{bi}z^t_b + \sum_{b\in[n]} A^*_{bi}\delta z^t_{b\to i})\\ &\quad - \mathcal{R}(A^*_{ai}z^t_a)\partial_1\eta^I(\sum_{b\in[n]} A^*_{bi}z^t_b + \sum_{b\in[n]} A^*_{bi}\delta z^t_{b\to i})\\ &\quad - \mathcal{I}(A^*_{ai}z^t_a)\partial_2\eta^I(\sum_{b\in[n]} A^*_{bi}z^t_b + \sum_{b\in[n]} A^*_{bi}\delta z^t_{b\to i}). \qquad (11)\end{aligned}$$

Also,

$$z_a^t = y_a - \sum_j A_{aj} x_j^t - \sum A_{aj} \delta x_{j\to a}^t. \tag{12}$$

By plugging (11) into (12), we obtain

$$\begin{aligned} -\sum_j A_{aj}\delta x_{j\to a}^t &= \sum_j A_{aj}\mathcal{R}(A_{aj}^* z_a^t)\partial_1\eta^R(\sum A_{bj}^* z_b^t + x_j^t) \\ &+ \sum_j A_{aj}\mathcal{I}(A_{aj}^* z_a^t)\partial_2\eta^R(\sum A_{bj}^* z_b^t + x_j^t) \\ &+ i\sum_j A_{aj}\mathcal{R}(A_{aj}^* z_a^t)\partial_1\eta^I(\sum A_{bj}^* z_b^t + x_j^t) \\ &+ i\sum_j A_{aj}\mathcal{I}(A_{aj}^* z_a^t)\partial_2\eta^I(\sum A_{bj}^* z_b^t + x_j^t), \end{aligned}$$

which completes the proof.

C Proof of Lemma 3.2

For notational simplicity, consider the case $m = 1$. The proof for the other values of m is exactly the same.

$$\begin{aligned} \Psi(1) &= \frac{1}{\delta}\mathbb{E}|\eta(X + Z_1 + iZ_2;\tau) - X|^2 \\ &= \frac{1-\epsilon}{\delta}\mathbb{E}|\eta(Z_1 + iZ_2;\tau)|^2 + \frac{\epsilon}{\delta}\mathbb{E}\left(\mathbb{E}_{\mu,\theta}|\eta(\mu e^{j\theta}Z_1 + iZ_2;\tau) - \mu e^{j\theta}|^2\right), \end{aligned} \tag{13}$$

where $\mathbb{E}_{\mu,\theta}$ denotes the conditional expectation given the variables μ, θ. The first term in (13) is independent of the phase and therefore we should prove that the second term is independent as well. Define

$$\Phi(\mu,\theta) \triangleq \mathbb{E}_{\mu,\theta}(|\eta(\mu e^{j\theta} + Z_1 + iZ_2;\tau) - \mu e^{j\theta}|^2), \tag{14}$$

We prove that Φ is independent of θ. Define $\mathbf{z} \triangleq (z_r, z_c)$, $d\mathbf{z} \triangleq dz_r dz_c$. We will use the following notation:

$$\begin{aligned} \alpha_z &\triangleq \sqrt{(\mu\cos\theta + z_r)^2 + (\mu\sin\theta + z_c)^2}, \\ c_r &\triangleq \frac{\mu\cos\theta}{\alpha_z}, \\ c_i &\triangleq \frac{\mu\sin\theta}{\alpha_z}. \end{aligned}$$

Define the two sets $A_\tau = \{(z_r, z_c) \mid \alpha_z < \tau\}$ and $A_\tau^c = \mathbb{R}^2 \backslash A_\tau$. We have

$$\Phi(\mu,\theta) = \int\limits_{\mathbf{z}\in A_\tau} \mu^2 \frac{1}{\pi} e^{-(z_r^2+z_c^2)} d\mathbf{z} + \int\limits_{\mathbf{z}\in A_\tau^c} |z_r + iz_c - \tau c_r - i\tau c_i|^2 \frac{1}{\pi} e^{-(z_r^2+z_c^2)} dz_r dz_c \tag{15}$$

Define $\beta \triangleq \mu\cos\theta + z_r$ and $\gamma \triangleq \mu\sin\theta + z_c$.

We obtain

$$\begin{aligned}
&\int\limits_{\mathbf{z}\in A_\tau^c} |z_r + jz_c - \tau c_r - j\tau c_i|^2 \frac{1}{\pi}\mathrm{e}^{-(z_r^2+z_c^2)} dz_r dz_c \\
&= \int\limits_{\sqrt{x^2+y^2}>\tau} \left|\beta - \mu\cos\theta + i(\gamma - \mu\sin\theta) - \frac{\tau\beta}{\sqrt{\beta^2+\gamma^2}} - i\frac{\tau\gamma}{\sqrt{\beta^2+\gamma^2}}\right|^2 \frac{1}{\pi}\mathrm{e}^{-(\beta-\mu\cos\theta)^2-(\gamma-\mu\sin\theta)^2} d\beta d\gamma \\
&= \int\limits_{\phi=0}^{2\pi}\int\limits_{r>\tau} |(r-\tau)\cos\phi - \mu\cos\theta + i((r-\tau)\sin\phi - \mu\sin\theta)|^2 \frac{1}{\pi}\mathrm{e}^{-(r\cos\phi-\mu\cos\theta)^2-(r\sin\phi-\mu\sin\theta)^2} r dr d\phi \\
&= \int\limits_{\phi=0}^{2\pi}\int\limits_{r>\tau} [(r-\tau)^2 + \mu^2 - 2\mu(r-\tau)\cos(\theta-\phi)]\mathrm{e}^{-r^2-\mu^2+2r\mu\cos(\theta-\phi)} r dr d\phi.
\end{aligned}$$

Periodicity of the cosine function proves that this term is also independent of the phase θ, and the proof is complete.

D Concavity of Ψ function

Lemma D.1. *The function $\Psi(m)$ is concave with respect to m.*

Proof. For the notational simplicity define $\sigma \triangleq \sqrt{m}$, $X_\sigma \triangleq \frac{X}{\sigma}$, and $A_\sigma \triangleq |X_\sigma - Z_1 + iZ_2|$. We have,

$$\begin{aligned}
\Psi(\sigma^2) &= \sigma^2\mathbb{E}\left(|\eta(X_\sigma + Z_1 + iZ_2;\tau) - X_\sigma|^2\right) = \\
&= \sigma^2\mathbb{E}\left(\mathbb{E}_X\left(|\eta(X_\sigma + Z_1 + iZ_2;\tau) - X_\sigma|^2\right)\right).
\end{aligned}$$

We first prove that $\Psi_X(\sigma^2) = \mathbb{E}_X\left(|\eta(X_\sigma + Z_1 + iZ_2;\tau) - X_\sigma|^2\right)$ is concave with respect to σ^2:

$$\begin{aligned}
\frac{d\Psi_X(\sigma^2)}{d\sigma^2} &= \mathbb{E}_X\left|\eta\left(X_\sigma + Z_1 + iZ_2;\tau\right) - X_\sigma\right|^2 + \frac{\sigma}{2}\frac{d}{d\sigma}\mathbb{E}_X\left|\eta\left(X_\sigma + Z_1 + iZ_2;\tau\right) - X_\sigma\right|^2 \\
&= \mathbb{E}_X\left|\eta\left(X_\sigma + Z_1 + iZ_2;\tau\right) - X_\sigma\right|^2 \\
&\quad -X_\sigma\mathbb{E}_X\left(\eta_1^R(X_\sigma + Z_1 + iZ_2;\tau) - 1\right)\left(\eta^R(X_\sigma + Z_1 + iZ_2;\tau) - X_\sigma\right).
\end{aligned}$$

It is therefore easy to see that

$$\begin{aligned}
\frac{d^2\Psi_X(\sigma^2)}{d^2\sigma^2} &= -\frac{X}{\sigma^3}\mathbb{E}_X\left(\eta_1^R(X_\sigma + Z_1 + iZ_2;\tau) - 1\right)\left(\eta^R(X_\sigma + Z_1 + iZ_2;\tau) - X_\sigma\right) \\
&\quad -\frac{X}{\sigma^3}\mathbb{E}_X\left(\eta_1^I(X_\sigma + Z_1 + iZ_2;\tau)\right)\left(\eta(X_\sigma + Z_1 + iZ_2;\tau)\right) \\
&\quad +\frac{X^2}{\sigma^4}\mathbb{E}_X\left(\eta_1^R(X_\sigma + Z_1 + iZ_2;\tau) - 1\right)^2 \\
&\quad +\frac{X^2}{\sigma^4}\mathbb{E}_X\left(\eta_1^I(X_\sigma + Z_1 + iZ_2;\tau)\right)^2. \qquad (16)
\end{aligned}$$

We have

$$\begin{aligned}
\eta_1^R(X_\sigma + Z_1 + Z_2; \tau) &= (1 - \frac{\tau Z_2^2}{A^3})\mathbb{I}(A > \tau), \\
\eta^R(X_\sigma + Z_1 + Z_2; \tau) &= (X_\sigma + Z_1 - \frac{\tau(X_\sigma + Z_1)}{A})\mathbb{I}(A \geq \tau), \\
\eta_1^I(X_\sigma + Z_1 + iZ_2; \tau) &= \frac{\tau(X_\sigma + Z_1)Z_2}{A^3}\mathbb{I}(A \geq \tau, \\
\eta^I(X_\sigma + Z_1 + iZ_2; \tau) &= Z_2 - \frac{\tau Z_2}{A}\mathbb{I}(A \geq \tau). \qquad (17)
\end{aligned}$$

Plugging (17) into (16) we have

$$\begin{aligned}
&\mathbb{E}_X\left(\eta_1^R(X_\sigma + Z_1 + iZ_2; \tau) - 1\right)\left(\eta^R(X_\sigma + Z_1 + iZ_2; \tau) - X_\sigma\right) \\
&+ \mathbb{E}_X\left(\eta_1^I(X_\sigma + Z_1 + iZ_2; \tau)\right)\left(\eta^I(X_\sigma + Z_1 + iZ_2; \tau)\right) = \frac{X}{\sigma}\mathbb{E}_X\left(\mathbb{I}(A \leq \tau) + \frac{\tau X Z_2^2}{\sigma A^3}\right). \qquad (18)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}_X\left(\eta_1^R(\frac{X}{\sigma} + Z_1 + iZ_2; \tau) - 1\right)^2 + \mathbb{E}_X\left(\eta_1^I(\frac{X}{\sigma} + Z_1 + iZ_2; \tau)\right)^2 \\
&= \mathbb{E}_X\left((1 - \frac{\tau Z_2^2}{A^3})\mathbb{I}(A \geq \tau) - 1\right)^2\left(\frac{\tau(X/\sigma + Z_1)Z_2}{A^3}\mathbb{I}(A \geq \tau)\right)^2 \\
&= E_X\left(\mathbb{I}(A \leq \tau) + \frac{\tau^2 Z_2^4}{A^6}\mathbb{I}(A \geq \tau) + \frac{\tau^2(\frac{X}{\sigma} + Z_1)^2 Z_2^2}{A^6}\mathbb{I}(A \geq \tau)\right). \qquad (19)
\end{aligned}$$

Combining (18) and (19) we obtain

$$\begin{aligned}
\frac{d^2\Psi_X(\sigma^2)}{d^2\sigma^2} &= \mathbb{E}_X(-\frac{\tau Z_2^2}{A^3} + \frac{\tau^2 Z_2^4}{A^6} + \frac{\tau^2(X/\sigma + Z_1)^2 Z_2^2}{A^6})\mathbb{I}(A \geq \tau) \\
&= \mathbb{E}_X(-\frac{\tau Z_2^2}{A^3} + \frac{\tau^2 A^2 Z_2^2}{A^6})\mathbb{I}(A \geq \tau) \leq 0. \qquad (20)
\end{aligned}$$

Finally, we use the fact that a convex combination of concave functions is concave to extend the concavity from $\Psi_X(m)$ to $\Psi(m)$, and the proof is complete. □

E Proof of Theorem 3.5

As proved in Lemma D.1, $\Psi(\sigma^2)$ is a concave function. Furthermore $\Psi(0) = 0$. Therefore 0 is a stable fixed point if and only if $\frac{d\Psi}{d\sigma^2}\big|_{m=0} < 1$. It is straightforward to calculate the derivative at zero and confirm that

$$\left.\frac{d\Psi}{dm}\right|_{m=0} = \frac{\rho\delta(1+\tau^2)}{\delta} + \frac{1-\rho\delta}{\delta}\mathbb{E}|\eta(Z_1 + iZ_2; \tau)|^2. \qquad (21)$$

Since $Z_1, Z_2 \sim N(0, 1/2)$ and are independent, the phase is uniform and the amplitude has Rayleigh distribution. Therefore, we have

$$\mathbb{E}|\eta(Z_1 + iZ_2; \tau)|^2 = 2\int_\tau^\infty \omega(\omega - \tau)^2 \mathrm{e}^{-\omega^2} d\omega. \qquad (22)$$

Plugging (22) in (21) and setting the derivative to 1 we obtain

$$\rho = \frac{\delta - 2\int_{\tau}^{\infty} \omega(\omega-\tau)^2 \mathrm{e}^{-\omega^2} d\omega}{\delta(1+\tau^2 - 2\int_{\tau}^{\infty} \omega(\omega-\tau)^2 \mathrm{e}^{-\omega^2} d\omega)}.$$

Therefore the optimal τ that achieves the highest phase transition satisfies

$$\left(-4\int_{\tau_*}^{\infty} \omega(\tau_*-\omega)\mathrm{e}^{-\omega^2} d\omega\right)\left(1+\tau_*^2 - 2\int_{\tau_*}^{\infty} \omega(\omega-\tau_*)^2 \mathrm{e}^{-\omega^2} d\omega\right) - \left(2\tau_* - 4\int_{\tau_*}^{\infty} \omega(\tau_*-\omega)\mathrm{e}^{-\omega^2} d\omega\right)\left(\delta - 2\int_{\tau_*}^{\infty} \omega(\omega-\tau_*)^2 \mathrm{e}^{-\omega^2}\right) = 0,$$

which in turn results in $\delta = \frac{4(1+\tau_*^2)\int_{\tau_*}^{\infty} \omega(\tau_*-\omega)\mathrm{e}^{-\omega^2} d\omega - 4\tau_* \int_{\tau_*}^{\infty} \omega(\omega-\tau_*)^2 \mathrm{e}^{-\omega^2 d\omega}}{-2\lambda + 4\int_{\tau_*}^{\infty} \omega(\tau_*-\omega)\mathrm{e}^{-\omega^2} d\omega}$. Plugging δ into the formula for ρ we obtain the formula in the theorem.

F Proof of Theorem 3.6

The proof is an application of Laplace's method. It is clear from Theorem 3.5 that as $\lambda \to \infty$, $\rho \to 0$ and $\delta \to 0$. Therefore, we should calculate the leading terms of ρ and δ. Using the Laplace's method, we can prove that

$$\int_{\lambda}^{\infty} \omega(\lambda-\omega)\mathrm{e}^{-\omega^2} d\omega \sim \frac{\mathrm{e}^{-\lambda^2}}{8\lambda^3}, \lambda \to \infty, \tag{23}$$

$$\int_{\lambda}^{\infty} \omega(\lambda-\omega)^2\mathrm{e}^{-\omega^2} d\omega \sim \frac{\mathrm{e}^{-\lambda^2}}{4\lambda^2}, \lambda \to \infty. \tag{24}$$

Plugging (23) and (24) into the formulas we have for ρ and δ in Theorem 3.5 we obtain

$$\begin{aligned} \delta &\sim \frac{\mathrm{e}^{-\lambda^2}}{2}, \lambda \to \infty, \\ \rho &\sim \frac{1}{\lambda^2}, \lambda \to \infty \end{aligned} \tag{25}$$

which completes the proof.

G Proof of Lemma 3.7

Since we have proved that the phase of the input signal does not affect the state evolution equation, we assume that the phase is equal to zero. Hence

$$r(\mu,\tau) = \mathbb{E}\,|\eta(\mu+Z_1+iZ_2;\tau)-\mu|^2 = \mathbb{E}(\eta^R(\mu+Z_1+iZ_2;\tau)-\mu)^2 + \mathbb{E}(\eta^I(\mu+Z_1+iZ_2;\tau))^2,$$

where $\eta^R(\mu+Z_1+iZ_2;\tau) = (\mu+Z_1-\frac{\tau(\mu+Z_1)}{A})\mathbb{I}(A\geq\tau)$, $\eta^I(\mu+Z_1+iZ_2;\tau) = (z_2-\frac{\tau Z_2}{A})\mathbb{I}(A\geq\tau)$ and $A=\sqrt{(\mu+Z_1)^2+Z_2^2}$. If we calculate the derivative of the risk function with respect to μ, we have

$$
\begin{aligned}
\frac{dr(\mu,\tau)}{d\mu} &= 2\mathbb{E}(\eta^R(\mu+Z_1+iZ_2;\tau)-\mu)(\frac{d\eta^R}{d\mu}-1)+2\eta^I(\mu+Z_1+iZ_2;\tau)\frac{d\eta^I}{d\mu} \\
&= \mathbb{E}(\eta^R(\mu+Z_1+iZ_2;\tau)-\mu)((1-\frac{\tau Z_2}{A^3})\mathbb{I}(A\geq\tau)-1)+\eta^I(\mu+Z_1+iZ_2;\tau)(\frac{\tau(\mu+Z_1)Z_2}{A^3})\mathbb{I}(A\geq\tau) \\
&= \mu\mathbb{E}(\mathbb{I}(A\leq\tau))-\mathbb{E}(Z_1-\frac{\tau(\mu+Z_1)}{A})(\frac{\tau Z_2^2}{A^3})+(Z_2-\frac{\tau Z_2}{A})\frac{\tau(\mu+Z_1)Z_2}{A^3} \\
&= \mu\mathbb{E}(\mathbb{I}(A\leq\tau))-\frac{\tau Z_1Z_2^2}{A^3}+\frac{\tau^2\mu Z_2^2}{A^4}+\frac{\tau^2Z_1Z_2^2}{A^4}+\frac{\tau\mu Z_2^2}{A^3}+\frac{\tau Z_2^2Z_1}{A^3}-\frac{\tau^2\mu Z_2^2}{A^4}-\frac{\tau^2Z_1Z_2}{A^4} \\
&= (\mu\mathbb{E}(\mathbb{I}(A\leq\tau)))+\mu\mathbb{E}\left(\frac{\tau Z_2^2}{A^3}\right)\geq 0.
\end{aligned}
$$

Therefore, the risk of the complex soft thresholding is an increasing function of μ. Furthermore,

$$
2\frac{dr(\mu,\tau)}{d\mu^2}=\frac{1}{\mu}\frac{dr(\mu,\tau)}{d\mu}=\mathbb{E}(\mathbb{I}(A\leq\tau))+\mathbb{E}\left(\frac{\tau Z_2^2}{A^3}\right).
$$

It is clear that the next derivative with respect to μ^2 is negative, and therefore the function is concave.

H Proof of Proposition 3.8

Let $\nu(\mu)=(1-\epsilon)\delta_0(\mu)+(1-\epsilon)G(\mu)$. We then have

$$
\begin{aligned}
\mathbb{E}|\eta(X+Z_1+iZ_2;\tau)-X|^2 &= (1-\epsilon)\mathbb{E}|\eta(Z_1+iZ_2;\tau)|^2+\epsilon\mathbb{E}_{X\sim G}\mathbb{E}_X|\eta(X+Z_1+iZ_2;\tau)-X|^2 \\
&= 2(1-\epsilon)\int_{w=\tau}^{\infty}w(w-\tau)^2\mathrm{e}^{-w^2}dw+\epsilon\mathbb{E}_{X\sim G}\mathbb{E}_X|\eta(X+Z_1+iZ_2;\tau)-X|^2.
\end{aligned}
\tag{26}
$$

Using Lemma 3.7 and Jensen inequality we prove that $\{G_m(\mu)\}_{m=1}^{\infty}$, $G_m(\mu)=\delta_m(\mu)$ is the least favorable sequence of distributions, i.e., for any distribution G

$$
\mathbb{E}_{X\sim G}\mathbb{E}_X|\eta(X+Z_1+iZ_2;\tau)-X|^2\leq\lim_{m\to\infty}\mathbb{E}_{X\sim G_m}\mathbb{E}_X|\eta(X+Z_1+iZ_2;\tau)-X|^2.
$$

Toward this goal we define $\tilde{G}(\mu)$ as $\delta_{\mu_0}(\mu)$, such that $\mu_0^2=E_G(X^2)$. In other words, $\tilde{G}$ and G have the same second moments. From the Jensen inequality we have

$$
\mathbb{E}_{X\sim G}\mathbb{E}_X|\eta(X+Z_1+iZ_2;\tau)-X|^2\leq\mathbb{E}_{X\sim\tilde{G}}\mathbb{E}_X|\eta(X+Z_1+iZ_2;\tau)-X|^2.
$$

Furthermore, from the monotonicity of the risk function proved in Lemma 3.7, we have

$$
\mathbb{E}_{X\sim\tilde{G}}\mathbb{E}_X|\eta(X+Z_1+iZ_2;\tau)-X|^2\leq\mathbb{E}_{X\sim G_m}\mathbb{E}_X|\eta(X+Z_1+iZ_2;\tau)-X|^2\ \ \forall m>\mu_0.
$$

Finally, the monotone convergence theorem indicates that

$$
\lim_{m\to\infty}\mathbb{E}_{X\sim G_m}\mathbb{E}_X|\eta(X+Z_1+iZ_2;\tau)-X|^2=1+\tau^2.
\tag{27}
$$

Combining (26) and (27) completes the proof.

References

[1] J. A. Tropp and S. J. Wright. Computational methods for sparse solution of linear inverse problems. *Proc. IEEE*, 98:948–958, 2010.

[2] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. on Sci. Computing*, 20:33–61, 1998.

[3] D. L. Donoho, A. Maleki, and A. Montanari. Message passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci.*, 106(45):18914–18919, 2009.

[4] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM J. on Sci. Computing*, 31(2):890–912, 2008.

[5] Z. Yang, C. Zhang, J. Deng, and W. Lu. Orthonormal expansion ℓ_1-minimization algorithms for compressed sensing. *Preprint*, 2010.

[6] A. Maleki and D. L. Donoho. Optimally tuned iterative thresholding algorithm for compressed sensing. *IEEE J. Select. Top. Signal Processing*, Apr. 2010.

[7] D. L. Donoho, A. Maleki, and A. Montanari. Noise sensitivity phase transition. *IEEE Trans. Inform. Theory*, 2010. submitted.

[8] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. Series B*, 58(1):pp. 267–288, 1996.

[9] A. Maleki and R. Baraniuk. Least favorable compressed sensing problems for the first order methods. *Proceedings of International Symposium on Information Theory*, 2011.

[10] M. Bayati and Andrea Montanri. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inform. Theory*, 57:764–785, 2011.

[11] G. Taubock and F. Hlawatsch. A compressed sensing technique for ofdm channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots. *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, 2008.

[12] J.S. Picard and A.J. Weiss. Direction finding of multiple emitters by spatial sparsity and linear programming. In *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, pages 1258 –1262, Sept. 2009.

[13] D. L. Donoho and J. Tanner. Precise undersampling theorems. *IEEE Trans. Inform. Theory.* submitted.

[14] D. L. Donoho and J. Tanner. Observed universality of phase transitions in high-dimensional geometry, with applications in modern signal processing and data analysis. *Philos. Trans. Roy. Soc. A*, 367(1906):4273–4293, 2009.

[15] M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Select. Top. Signal Processing*, 1(4):586–598, 2007.

[16] S. Wright, R. Nowak, and M. Figueiredo. Sparse reconstruction by separable approximation. *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, 2009.

[17] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. A method for large-scale l1-regularized least squares. *IEEE J. Select. Top. Signal Processing*, 1(4):606–617, Dec. 2007.

[18] J. Huang and T. Zhang. The benefit of group sparsity. *Preprint arXiv:0901.2962*, 2009.

[19] J. Peng, J. Zhu, A. Bergamaschi, W. Han, D.Y. Noh, J.R. Pollack, and P. Wang. Regularized multivariate regression for identifying master predictors with application to integrative genomics study of breast cancer. *Ann. Appl. Stat.*, 4(1):53–77, 2010.

[20] M.F. Duarte, W.U. Bajwa, and R. Calderbank. The performance of group lasso for linear regression of grouped variables. Technical report, Technical Report TR-2010-10, Duke University, Dept. Computer Science, Durham, NC, 2011.

[21] E. Van Den Berg and M.P. Friedlander. Theoretical and empirical results for recovery from multiple measurements. *IEEE Trans. Inform. Theory*, 56(5):2516–2527, 2010.

[22] J. Chen and X. Huo. Theoretical results on sparse representations of multiple-measurement vectors. *IEEE Trans. Signal Processing*, 54(12):4634–4643, 2006.

[23] Y.C. Eldar, P. Kuppinger, and H. Bolcskei. Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Trans. Signal Processing*, 58(6):3042–3054, 2010.

[24] X. Lv, G. Bi, and C. Wan. The group lasso for stable recovery of block-sparse signal representations. *IEEE Trans. Signal Processing*, 59(4):1371–1382, 2011.

[25] M. Stojnic, F. Parvaresh, and B. Hassibi. On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Trans. Signal Processing*, 57(8):3075–3085, 2009.

[26] M. Stojnic. ℓ_2/ℓ_1-optimization in block-sparse compressed sensing and its strong thresholds. *IEEE J. Select. Top. Signal Processing*, 4(2):350–357, 2010.

[27] M. Stojnic. Block-length dependent thresholds in block-sparse compressed sensing. *Preprint arXiv:0907.3679*, 2009.

[28] S. Ji, D. Dunson, and L. Carin. Multi-task compressive sensing. *IEEE Trans. Signal Processing*, 57(1):92–106, 2009.

[29] S. Bakin. Adaptive regression and model selection in data mining problems. *PhD Thesis, Australian National University*, 1999.

[30] L. Meier, S. Van De Geer, and P. Buhlmann. The group lasso for logistic regression. *J. Roy. Statist. Soc. Ser. B*, 70(Part 1):53–71, 2008.

[31] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *J. Roy. Statist. Soc. Ser. B*, 68(1):49–67, 2006.

[32] F. R. Bach. Consistency of the group lasso and multiple kernel learning. *J. Machine Learning Research*, 9:1179–1225, Jun. 2008.

[33] Y. Nardi and A. Rinaldo. On the asymptotic properties of the group lasso estimator for linear models. *Electron. J. Statist.*, 2:605–633, 2008.

[34] K. Lounici, M. Pontil, A. B. Tsybakov, and S. van de Geer. Taking advantage of sparsity in multi-task learning. *arXiv:0903.1468v1*, 2009.

[35] D. Malioutov, M. Cetin, and A.S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Processing*, 53(8):3010 – 3022, Aug. 2005.

[36] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26(3):301–321, 2008.

[37] T. Blumensath and M. E. Davies. How to use the iterative hard thresholding algorithm. *Proc. Work. Struc. Parc. Rep. Adap. Signaux (SPARS)*, 2009.

[38] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.*, 27(3):265–274, 2009.

[39] D. L. Donoho, I. Drori, Y. Tsaig, and J. L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. *Stanford Statistics Department Technical Report*, 2006.

[40] A. Maleki. Approximate message passing algorithm for compressed sensing. *Stanford University PhD Thesis*, 2010.

[41] D. L. Donoho, A. Maleki, and A. Montanari. Construction of message passing algorithms for compressed sensing. *Preprint*, 2010.

[42] M. J. Wainwright G. Obozinski and M. I. Jordan. Support union recovery in high-dimensional multivariate regression. *Ann. Stat.*, 39:1–47, 2011.

[43] D. L. Donoho and Andrea Montanari. Simultaneous sparse recovery by vector approximate message passing (vamp).

[44] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Inform. Theory*, 56(4):1982 –2001, Apr. 2010.

[45] M. Bayati and A. Montanari. The lasso risk for gaussian matrices. *arXiv:1008.2581v1*, 2011.

[46] Z. Yang and C. Zhang. Sparsity-undersampling tradeoff of compressed sensing in the complex domain. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, 2011.

[47] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM J. Multiscale Model. Simul.*, 4(4):1168–1200, 2005.